MrPaw

Ryan Giordano (`rgiordano@berkeley.edu`, UC Berkeley), ..,

Berkeley BSTARS September 2025

Are US non-voters becoming more Republican?

Blue Rose research says yes:

"Politically disengaged voters have become much more Republican, And because less-engaged voters swung away from [Democrats], an expanded electorate meant a more Republican electorate."

[Blue Rose Research, 2024] (On Ezra Klein show, major professional pollsters)

***On Data and Democracy* says no:**

"Claims of a decisive pro-Republican shift among the overall non-voting population are not supported by the most reliable, large-scale post-election data currently available."

[Bonica et al., 2025] (Berkeley professor co–author, major professional researchers)

- The problem is very hard (it's difficult to accurately poll non–voters)
- Different data sources
- **Very different statistical methods:** $\star$
  - Blue Rose uses Bayesian hierarchical modeling (MrP)
  - The CES uses weighted averages (calibration weighting)

Our contribution

We provide a calibration weighting interpretation of MrP analyses that:

- Is easily computable from MCMC draws and standard software, and
- Defines MrP versions of the diagnostics that motivate calibration weighting.

Our "MrP approximate weights" (MrPaw) admit apples-to-apples comparisons between these very different methodologies.

Outline

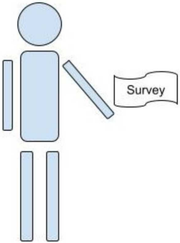
- Introduce the statisical problem and two methods (calibration weighting and MrP)
- Describe one of the classical diagnostics (covariate balance)
- Define MrPaw & state a key theorem
- Real-world results
- Future directions

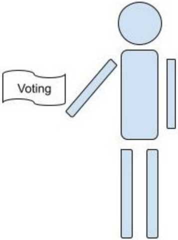
The basic problem

We have a survey population, for whom we observe:

- Covariates **x** (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to "do you support policy such–and–such")

We want the average response in a target population, in which we observe only covariates.





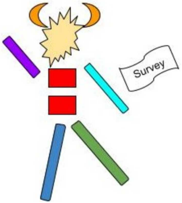
Observe $(\mathbf{x}_s, y_s)$ for $s = 1, \ldots, \mathcal{N}_S$ | Observe $\mathbf{x}_p$ for $p = 1, \ldots, \mathcal{N}_T$

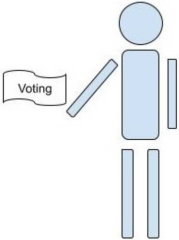
The basic problem

We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to "do you support policy such–and–such")

We want the average response in a target population, in which we observe only covariates.





Observe $(\mathbf{x}_s, y_s)$ for $s = 1, \ldots, \mathcal{N}_S$ | Observe $\mathbf{x}_p$ for $p = 1, \ldots, \mathcal{N}_T$

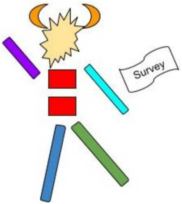
The problem is that the populations are very different.

The basic problem

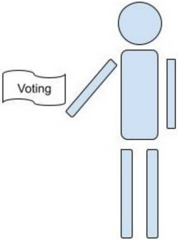
We have a survey population, for whom we observe:

- Covariates $\mathbf{x}$ (e.g. race, gender, zip code, age, education level)
- Responses y (e.g. A binary response to "do you support policy such–and–such")

We want the average response in a target population, in which we observe only covariates.



Observe $(\mathbf{x}_s, y_s)$ for $s = 1, \ldots, \mathcal{N}_S$



Observe $\mathbf{x}_p$ for $p = 1, \ldots, \mathcal{N}_T$

The problem is that the populations are very different.

Our survey results may be biased.

How can we use the covariates to say something about the target responses?

Two approaches

We want $\mu := \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} y_p$, but don't observe y_p in the target population.

- Assume $p(y|x)$ is the same in both populations,
- But the distribution of x may be different in the survey and target.

Calibration weighting	**Bayesian hierarchical modeling (MrP)**
Choose "calibration weights" w_s (e.g. raking weights)	Choose a model $\mathcal{P}(y|x,\theta)$ and prior $\mathcal{P}(\theta)$ (e.g. Hierarchical logitstic regression)
$\hat{\mu}_{\text{CAL}} = \frac{1}{N_S} \sum_{n \in \mathcal{N}_S} w_s y_s$	Take $\hat{y}_p = \mathbb{E}_{\mathcal{P}(\theta \vert \text{Survey data})} \left[y \vert \mathbf{x}_p \right]$ and $\hat{\mu}_{\text{MRP}} = \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} \hat{y}_p$
Dependence on y_s is obvious (w_s typically chosen using only **x**)	Dependence on y_s very complicated (Typically via MCMC draws from $\mathcal{P}(\theta \vert \text{Survey data})$)
Weights give interpretable diagnostics: • Frequentist variability • Partial pooling • Regressor balance	**Black box**

We open the MrP black box, and provide versions of all these diagnostics, for nonlinear hierarchical models fit with MCMC.

Covariate balance

What do we want out of calibration weights?

$$\text{Target average} = \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} y_p \approx \frac{1}{N_S} \sum_{n \in \mathcal{N}_S} w_s y_s = \text{Weighted survey average}$$

We can't check this, because we don't observe y_p. But we can check whether

$$\frac{1}{N_T} \sum_{n \in \mathcal{N}_T} \mathbf{x}_p = \frac{1}{N_S} \sum_{n \in \mathcal{N}_S} w_s \mathbf{x}_s$$

Such weights satisfy "covariate balance" for $\mathbf{x}$.

You can check covariate balance for any calibration weighting estimator.

Even more, covariate balance is the criterion for a popular class of calibration weight estimators:

Raking calibration weights

"Raking" selects weights that

- Are as "close as possible" to some reference weights
- Under the constraint that they balance some selected regressors.

Why covariate balance?

Why covariate balance?

We want to balance $f(\mathbf{x})$ because we think $\mathbb{E}\left[y|\mathbf{x}\right]$ might plausibly vary $\propto f(\mathbf{x})$, and want to check whether our estimator can capture this variability.

This motivates the following **generalized covariate balance check**:

General covariate balance check (informal)

Define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the change this is supposed to induce in the target population.

Covariate balance checks whether our estimators produce the same change.

Covariate balance

General covariate balance check (formal)

Pick a small δ, and define a *new response variable* $\tilde{y}$ such that

$$\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x}).$$

We know the expected change this perturbation produces in the target distribution:

$$\mathbb{E}\left[\mu(\tilde{y}) - \mu(y)|\mathbf{x}\right] = \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} \left(\mathbb{E}\left[\tilde{y}|\mathbf{x}_p\right] - \mathbb{E}\left[y|\mathbf{x}_p\right]\right) = \delta \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} f(\mathbf{x}_p)$$

Covariate balance checks whether an estimator $\hat{\mu}(\cdot)$ produces the same change:

$$\hat{\mu}(\tilde{y}) - \hat{\mu}(y) \overset{\text{check}}{=} \delta \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} f(\mathbf{x}_p).$$

When $\hat{\mu}(y)$ is a calibration estimator, this is the same as covariate balance in expectation:

$$\mathbb{E}\left[\hat{\mu}(\tilde{y}) - \hat{\mu}(y)|\mathbf{x}\right] = \delta \frac{1}{N_S} \sum_{n \in \mathcal{N}_S} w_s f(\mathbf{x}_p) \overset{\text{check}}{=} \delta \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} f(\mathbf{x}_p).$$

But now all we need to do is compare $\hat{\mu}(\tilde{y}) - \hat{\mu}(y)$ for “nearby” $\tilde{y}$ and y.

We need to approximate $\hat{\mu}_{\text{MRP}}(\tilde{y}) - \hat{\mu}_{\text{MRP}}(y)$.

Step one: Define weights.

Noting that $w_s = \frac{d}{dy_s}\hat{\mu}_{\text{CAL}}$, we can define

$$w_s^{\text{MRP}} := \frac{d}{dy_s}\hat{\mu}_{\text{MRP}}.$$

It happens that the needed derivatives are given by simple a posterior covariances involving only the inverse link function $m(\mathbf{x};\theta)$ and log likelihood [Giordano et al., 2018]:

$$\frac{d\hat{y}_p}{dy_s} = \text{Cov}_{\mathcal{P}(\theta|\text{Survey data})}\left(m(\mathbf{x}_p;\theta), \frac{\partial}{\partial y}\log p(y|\theta,\mathbf{x}_s)\right)$$

These can be computed using standard MCMC software [Bürkner, 2017].

No other weight definition will do — in some cases, MrP is exactly a calibration estimator (e.g. linear regression with flat priors), and we want the definitions to coincide in that case.

Covariate balance

How to form a notion of covariate balance for estimators that are not weighted averages?

Calibration weights

$\hat{\mu}_{\text{CAL}} = \frac{1}{N_S} \sum_{n \in \mathcal{N}_S} w_s y_s$

MrP

Take $\hat{y}_p = \mathbb{E}_{\mathcal{P}(\theta | \text{Survey data})} \left[y | \mathbf{x}_p \right]$ and

$\hat{\mu}_{\text{MRP}} = \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} \hat{y}_p$

Step two: Specify a Taylor series.

Suppose we wanted to re–compute MrP with new survey responses y_s^{new}.

$$\hat{\mu}_{\text{MRP}}(y_1^{\text{new}}, \ldots, y_{\mathcal{N}_S}^{\text{new}}) = \sum_{n \in \mathcal{N}_S} w_s^{\text{MRP}} (y_s^{\text{new}} - y_s) + \text{Residual}$$

In general, MrP is truly nonlinear. The residual is only small when $y_s^{\text{new}} \approx y_s$!

Covariate balance

Step three: Define a data perturbation that captures regression balance.

Recall that our y is binary. How can we produce a $\tilde{y}$ such that $\mathbb{E}\left[\tilde{y}|\mathbf{x}\right] = \mathbb{E}\left[y|\mathbf{x}\right] + \delta f(\mathbf{x})$?

- Use an estimate of $\mathbb{E}\left[y|\mathbf{x}\right]$ to draw new binary data, or
- Allow $\tilde{y}$ to take values other than $\{0, 1\}$ and set $\tilde{y} = y + \delta f(\mathbf{x})$.

Focus on the second (we have examples of the first).

Covariate balance

Theorem

- Let $\tilde{y} = y + \delta f(\mathbf{x})$,
- $\hat{\mu}_{\mathrm{MRP}}$ be a hierachical logistic regression posterior expectation, and
- $\mathcal{F}$ be a class of uniformly bounded functions on $\mathbf{x}$.

Then, with probability approaching one, as $N \to \infty$,

$$\sup_{f \in \mathcal{F}} \left(\hat{\mu}_{\mathrm{MRP}}(\tilde{y}) - \left(\hat{\mu}_{\mathrm{MRP}}(y) + \sum_{n \in \mathcal{N}_S} w_s^{\mathrm{MRP}} \delta f(\mathbf{x}_s) \right) \right) = O(\delta^2) \quad \text{as } \delta \to 0$$

The supremum over $\mathcal{F}$ is the primary technical contribution! It means we are justified in searching over regressors to find imbalance.

Draws on our prior work on uniform and finite–sample error bounds for Bernstein–von Mises theorem–like results [Giordano and Broderick, 2024, Kasprzak et al., 2025].

In practice, compute

$$\mathrm{Imbalance}(f) := \sum_{n \in \mathcal{N}_S} w_s^{\mathrm{MRP}} f(\mathbf{x}_s) - \frac{1}{N_T} \sum_{n \in \mathcal{N}_T} f(\mathbf{x}_p)$$

for any $f(\cdot)$ you think might capture variability in y.

References

Blue Rose Research. 2024 Election Retrospective Presentation. https://data.blueroseresearch.org/2024retro-download, 2024. Accessed on 2024-10-26.

A. Bonica, R. Fordham, J. Grumbach, and E. Tiburcio. Did non-voters really flip Republican in 2024? The evidence says no. https://data4democracy.substack.com/p/did-non-voters-really-flip-republican, April 2025.

Paul-Christian Bürkner. brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, 80(1): 1–28, 2017. doi: 10.18637/jss.v080.i01.

R. Giordano and T. Broderick. The Bayesian infinitesimal jackknife for variance, 2024. URL https://arxiv.org/abs/2305.06466.

R. Giordano, T. Broderick, and M. I. Jordan. Covariances, robustness and variational bayes. *Journal of machine learning research*, 19(51), 2018.

M. Kasprzak, R. Giordano, and T. Broderick. How good is your Laplace approximation of the bayesian posterior? Finite-sample computable error bounds for a variety of useful divergences, 2025. URL https://arxiv.org/abs/2209.14992.